AD_____

Award Number: DAMD17-99-1-9119

TITLE: A New Model for Estimation of Breast Cancer Risk

PRINCIPAL INVESTIGATOR: Maryellen L. Giger, Ph.D.

CONTRACTING ORGANIZATION:

The University of Chicago
Chicago, Illinois 60637

REPORT DATE: July 2002

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 2002	Final (14 Jun 99 - 13 Jun 02)

4. TITLE AND SUBTITLE
A New Model for Estimation of Breast Cancer Risk

5. FUNDING NUMBERS
DAMD17-99-1-9119

6. AUTHOR(S)
Maryellen L. Giger, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The University of Chicago
Chicago, Illinois 60637

E-Mail: m-giger@uchciago.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 Words)*

Cancer risk is the probability that cancer will occur in a given population. Research on cancer risk seeks to identify populations with a high probability of developing cancer. The goal of this research is to merge a computerized analysis of mammograms, which characterizes the breast pattern, with information of a woman's personal and family histories into a novel model for use in estimating risk of breast cancer. We have shown that computer-extracted features of mammographic parenchymal patterns can be used in the prediction of breast cancer risk. This has been demonstrated using three approaches: (1) correlation with clinical models of Gail and Claus, (2) separation between women at low risk and those with a positive gene testing result, and (3) separation between women at low risk and those that have breast cancer. In addition, we have shown, that the inclusion of the mammographic features with age increase the predictive power over the use of age alone in the prediction of breast cancer risk. We have also shown that with our method, the performance of the features and the classifier are quite dependent on ROI location within the breast and only slightly dependent on ROI size.

14. SUBJECT TERMS	15. NUMBER OF PAGES
breast cancer, computerized cancer risk assessment, texture-based image analysis	17
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

Final Report
Proposal Title: A New Model for the Estimation of Breast Cancer Risk

P.I.: Maryellen L. Giger, Ph.D.

INTRODUCTION:

Cancer risk is the probability that cancer will occur in a given population. Research on cancer risk seeks to identify populations with a high probability of developing cancer. The goal of this research is to merge a computerized analysis of mammograms, which characterizes the breast pattern, with information of a woman's personal and family histories into a novel model for use in estimating risk of breast cancer. The specific aims include 1. Creating a database of mammograms, along with tabulated clinical information of women at low risk and high risk for breast cancer; 2. Developing a new model using computer methods for merging mammographic information with clinical information; and 3. Evaluating the efficacies of the new model compared to currently used methods of risk assessment. The main hypothesis to be tested is that given a group of women, the new computerized risk model that merges computerized analyses of mammograms with clinical information should yield a novel way for identifying those women at risk for breast cancer. Potential uses of this innovative model include 1) serving as a means to assess the cancer risk of women undergoing routine screening mammography and thus, identifying those women that may require closer scrutiny and 2) serving as a means to monitor the cancer risk of women undergoing chemoprevention treatments. The research is novel in that currently there does not exist a reliable means to assess the cancer risk of individual women using both mammographic and clinical information. In addition, if a woman knew that she was at an increased risk of breast cancer, it is likely that she would better comply with screening mammography programs. In the future, a successful model could also be used to assess the effect of chemoprevention on a women's parenchymal pattern and thereby, overall risk.

BODY:

Task 1. Establishment of database (mos. 1-30)

The high-risk database was collected within the University of Chicago Cancer Risk Clinic and consists of mammograms, pedigree information, epidemiological data and related biological specimens from patients with a family history of breast cancer. All

mammograms done since1990 were considered for collection for all participants irrespective of their cancer status. For each, breast cancer risk assessment was performed using both Gail and Claus models and genetic testing whenever possible. A low-risk database was also collected from our breast cancer screening program and includes mammograms and clinical information on women undergoing routine screening mammograms. The low risk database was developed to include women who are age-matched to reflect the age of women in our high risk database. We collected 380 cases (yielding over 1000 films), which includes 143 "low risk" cases, 222 high/moderate risk cases, and 35 BRCA1/BRCA2-mutation carriers. The low risk and high/moderate cases were deemed to be low or moderate/high risk by the use of the clinical Gail and Claus models by the University of Chicago Cancer Risk Clinic. In addition, the clinical information of age for each patient was tabulated. The mammograms are converted to digital format by using a laser film scanner (2048 by 2048 matrix with 12-bit quantization). Such high spatial resolution is necessary in order to adequately retain the high-frequency texture patterns.

Task 2. Development of risk model including mammographic markers and clinical information (mos. 3-30)

Computerized analysis of the parenchymal pattern is based on various texture analysis methods we have developed in our laboratory including Fourier spectra analysis, histogram analysis, and artificial neural networks. Fourteen image features were extracted within the regions of each digitized mammogram. These features can be grouped into (i) features based on the absolute values of the gray levels, (ii) features based on gray-level histogram analysis, (iii) features based on the Fourier transform, and (iv) features based on the spatial relationship among gray levels.

We employed three different approaches to relate these mammographic features to breast cancer risk. In one approach, the features were used to distinguish mammographic patterns seen in low-risk women from those who inherited a mutated form of the *BRCA1/BRCA2* gene. In another approach, the features were related to risk as determined from existing clinical models (*Gail* and *Claus* models). Stepwise linear discriminant analysis was employed to identify features that were useful in differentiating between "low-risk" women and *BRCA1/BRCA2*-mutation carriers. Stepwise linear regression analysis was employed to identify useful features in predicting the risk as estimated from the *Gail* and *Claus* models. In the third approach, the features were used to characterize mammographic patterns seen in low-risk women and in women who have breast cancer. Stepwise linear logistic regression was employed to identify useful features to differentiate between the mammographic patterns of low-risk women and

women with breast cancer. The relationship between the image patterns and the risk of developing breast cancer was identified based on the odds ratios associated with these image features. The computer-extracted mammographic features identified from these three approaches were similar. The results from these studies show that women who have dense breasts and whose mammographic patterns are coarse and low in contrast have an increased risk of developing breast cancer. The consensus of the findings from the three different approaches substantiated the existing results. (Presented CARS 2000) Futher investigation of the gene carrier group resulted in a RSNA 2000 presentation (November, 2000) and an "in-press" Radiology paper.

We also analyzed the contributions of age and computer-extracted mammographic features in the prediction of breast cancer risk. We assessed the contribution of the computer-extracted features to risk prediction in terms of percent increase in the prediction power (r^2) when age (the single most important risk factor for breast cancer) was used alone and when the mammographic features were included. The inclusion of the mammographic features increased the prediction power (r^2) from 0.08 and 0.16 (age alone) to 0.17 and 0.32, yielding an increase of 113% an d 100% in r^2 for predicting the risk as estimated from the Gail and Claus models. The substantial increase in r^2 indicates the important contribution of these mammographic features in risk prediction and the need to incorporate in predicting breast cancer risk. (Presented IWDM 2000)

Task 3. Evaluation methods (mos. 20-36)

Correlation analysis was used in evaluating the performance of the computer-extracted features and the clinical features. Linear correlation analysis was performed to determine the correlation among the output of the new model and the Gail risk model (or Claus model). We used the combined model based on the first two models (gene mutation vs. low-risk and with cancer vs. without cancer) and evaluated the performance of the combined measures using the Gail model.

We have entered into a collaborative agreement with the University of Toronto to analyze data from the Ontario Breast Screening Program including 400 case control pairs. In a nested case-control database, the cases will correspond to women who will have developed cancer and the control will correspond to women who will have stayed cancer free during the period. We will calculate the clinical markers (e.g., Gail) and the mammographic features of the initial examination prior to the 5 to 8 year follow-up. Multivariate analysis will be used to examine the relationship between the new model and risk of breast cancer while controlling for other risk factors such as age at menarche

and parity. A proportional-hazards regression model will be used to calculate the relative risk for each radiographic marker.

In preparation for this analysis, we investigated the effect of ROI size on the computer-extracted parenchymal texture features. The results showed that the ability of the texture features to discriminate between high risk (BRCA1/BRCA2 mutation carriers) and low risk women was dependent on ROI location but only slightly dependent on ROI size. This work was presented at the 2002 AAPM meeting.

KEY RESEARCH ACCOMPLISHMENTS:

- Increased our database of high and low risk cases, especially those with positive BRCA1/BRCA2 testing.
- Verified the texture features for characterizing the breast parenchyma using three different approaches -- all yielding the same result
- Performed initial study looking at the contribution of age and mammographic features to breast cancer risk prediction
- Evaluated computer-extracted texture features with respect to ROI location and ROI size

REPORTABLE OUTCOMES:

1. Huo Z, Giger ML, Olopade OI: Analysis of the relative contributions of mammographic features and age to breast cancer risk prediction. Zhimin Huo, Maryellen L. Giger and Olufunmilayo I. Olopade, **Presentation** at International Workshop on Digital Mammography 2000 (Toronto, Canada)
2. Huo Z, Giger ML: Incorporation of clinical data into a computerized method for the assessment of mammographic breast lesions. **Proceeding Paper** Proc. SPIE 2000, 3979:148-152, 2000.
3. Huo Z, Giger ML, Olopade OI: Computerized analysis of mammographic patterns of women with and without breast cancer. Zhimin Huo, Maryellen L. Giger and Olufunmilayo I. Olopade, **Presentation** at CARS 2000 (San Fransico, CA)

4. Huo Z, Giger ML, Wolverton DE, Zhong W, Cummings S, Olopade OI: Computerized analysis of mammographic parenchymal patterns for breast cancer risk assessment: Feature selection. **Journal Article** Medical Physics 27:4-12, 2000.
5. Huo Z, Giger ML, Zhong W, Nishikawa, RE, Wolverton DE, Olopade OI: Mammographic parenchymal patterns as predictors for breast cancer risk. **Presentation** at 86th Scientific Assembly and Annual Meeting of Radiological Society of North America, Chicago, Illinois, 2000.
6. Huo Z, Giger ML, Zhong W, Olopade OI: Analysis of relative contributions of mammographic features and age to breast cancer risk prediction. **Proceeding Paper** Digital Mammography 2000, Proc. 5th International Workshop on Digital Mammography, Medical Physics Publishing, Wisconsin pp. 732-736, 2001.
7. Huo Z, Giger ML, Olopade OI, Wolverton DE, Weber BL, Metz CE, Cummings S, Zhong W: Computerized analysis of digitized mammograms of BRCA1/BRCA2 gene mutation carriers. **Journal article** Radiology (in press), 2002.
8. Li Hui, Giger ML, Huo Z, Olopade O, Lan L, Bonta I: Computerized analysis of mammographic patterns for assessing breast cancer risk: Effect of ROI size and location. **Presentation** at 2002 AAPM meeting in Montreal, Canada , 2002.
9. Giger ML has a **RO1 grant being submitted** to NCI, which formally includes the collaboration with the University of Toronto and the Ontario Breast Cancer Screening Program (made possible by the results from the army idea grant).

PERSONNEL WHO RECEIVED SUPPORT FROM THE GRANT

Maryellen Giger, Ph.D.
Funmi Olopade, M.D.
Dulcy Wolverton, M.D
Zhimin Huo, Ph.D.
Weiming Zhong, M.S.
Hui Li, Ph.D.
Chun Wai Chan, M.S.
Michael Carlin, R. T.

CONCLUSIONS:

We have shown that computer-extracted features of mammographic parenchymal patterns can be used in the prediction of breast cancer risk. This has been demonstrated (on the developing database) using three approaches: (1) correlation with clinical models of Gail and Claus, (2) separation between women at low risk and those with a positive gene testing result, and (3) separation between women at low risk and those that have breast cancer. In addition, we have shown, that the inclusion of the mammographic features with age increase the predictive power over the use of age alone in the prediction of breast cancer risk. We have also shown that with our method, the performance of the features and the classifier are quite dependent on ROI location within the breast and only slightly dependent on ROI size.

REFERENCES

1. Gail MH, Benichou J. Assesing the risk of breast cancer in individuals. In Cancer Prevention edited by Saul Rosenberg.Published by JB Lippincott Company. 1-15.1992.

2. Kosary CL, Ries LAG, Miller BA, Harris A and Edwards BK. SEER cancer statistics review, 1973-1992: tables and graphs. Bethesda, MD, National Cancer Institute, 1995

3. Claus EB, Risch NJ, Thompson WD: Genetic analysis of breast cancer in the Cancer in the Cancer and Steroid hormone study. Am J Hum Genet. 48: 232-242, 1991.

4. Offit K and Brown K. Quantitation of familial cancer risk: a resource for clinical oncologists. J Clin Oncol 1994; 86:620-625.

5. Boyd NF, O'Sullivan B, Fishell E et al.: Mammographic patterns and breast cancer risk: methodological standards and contradictory results. J Natl Cancer Inst 72:1253-1259, 1984.

6. Wolfe JN: Breast patterns as an index of risk for developing breast cancer. AJR 126: 1130-1139, 1976.

7. Brisson J, Morrison AS and Khalid N. Mammographic parenchymal features and breast cancer in the Breast Cancer Detection Demonstration Project. J Natl Cancer Inst 1980; 80:1534-1540.

8. Saftlas AF, Hoover RN, Brinton LA, Szklo M, Olson DR, Salane M and Wolfe JN. Mammographic densities and risk of breast cancer. Cancer 1991; 67:2833-2838.

9. Byrne C, Schairer C, Wolfe J, Parekh N, Salane M, Brinton LA, Hoover R and Haile R. Mammographic features and breast cancer risk: effects with time, age, and menopause status. J Natl Cancer Inst 1995; 87:1622-1629.

10. Egan RL, Mosteller RC: Breast cancer mammography patterns. Cancer 40: 2087-2090, 1977.

11. Oza AM, Boyd NF: Mammographic parenchymal patterns: a marker of breast cancer risk. Epidemiologic Rev. 15:196-208, 1993.

12. Boyd NF, Jensen HM, Cooke G, et al.: Relationship between mammographic and histological risk factors for breast cancer. J Natl Cancer Inst 84: 1170-1179, 1992.

13. Giger ML: "Future of Breast Imaging. Computer-Aided Diagnosis". In: AAPM/RSNA Categorical Course on the Tech Aspects of Breast Imaging, 3rd ed, (Haus A. and Yaffe M., eds.) pp. 287-302, 1994.

14. Vyborny CJ, Giger ML: Computer vision and artificial intelligence in mammography. AJR 162: 699-708, 1994.

15. Magnin IE, Cluzeau F, Odet CL: Mammographic texture analysis: an evaluation of risk for developing breast cancer. Optical Engineering.25:780-784. 1986.

16. Caldwell CB, Stapleton SJ, Holdsworth DW, Jong RA, Weiser WJ, Cooke G, Yaffe MJ: Characterization of mammographic parenchymal pattern by fractal dimension. Phys. Med. Biol. 35:235-247. 1990.

17. Taylor P, Hajnal S, Dilhuydy M-H, Barreau B: Measuring image texture to separate "difficult" from "easy" mammograms. British J Rad 67: 456-463, 1994.

18. Tahoces PG, Correa J, Souto M, et al.: Computer-assisted diagnosis: the classification of mammographic breast parenchymal patterns. Phys Med Biol, 40: 103-117, 1995.

19. Byng JW, Boyd NF, Fishell E, Jong RA, Yaffe MJ: The quantitative analysis of mamographic densities. Phys Med Biol 39: 1629-1638, 1994.

20. Byng JW, Boyd NF, Fishell E, Jong R and Yaffe MJ. Automated analysis of mammographic densities. Phys Med Biol 1996; 1996:909-923.

21. Byng JW, Yaffe MJ, Lockwood G, et al.: Automated analysis of mammographic densities and breast cancinoma risk. Cancer 80: 66-74, 1997.

22. Giger ML, Doi K, MacMahon H, Nishikawa RM, Hofmann KR, et al.: An "intelligent" workstation for computer-aided diagnosis". RadioGraphics 13: 647-656, 1993.

23. Nishikawa RM, Haldemann RC, Papaioannou J, Giger ML, Lu P, Schmidt RA, Wolverton DE, Bick U, Doi K: Initial experience with a prototype clinical "intelligent" mammography workstation for computer-aided diagnosis. Proc SPIE 2434: 65-71, 1995.

24. Giger ML, Nishikawa RM, Kupinski MA, Bick U, Zhang M, Schmidt RA, et al.: Computerized detection of breast lesions in digitized mammograms and results with a clinically-implemented intelligent workstation. CAR'97 pgs. 325-330, 1997, 1997.

25. Huo Z, Giger ML, Vyborny CJ, Bick U, Lu P, Wolverton DE, Schmidt RA: Analysis of spiculation in the computerized classification of mammographic masses" Medical Physics, 1995.

26. Huo Z, Giger ML, Vyborny CJ, Wolverton DE, Schmidt RA, Doi K: Automated computerized classification of malignant and benign mass lesions on digitized mammograms. Academic Radiology 5: 155-168, 1998.

27. Jiang Y, Nishikawa RM, Wolverton DE, Metz CE, Giger ML, Schmidt RA, Vyborny CJ, Doi K: Automated feature analysis and classification of malignant and benign clustered microcalcifications. Radiology 198:671-678, 1996.

28. Huo Z, Giger ML, Olopade OI, et al: Computer-aided diagnosis: Breast cancer risk assessment from mammographic parenchymal patterns in digital mammograms. Digital Mammography '96. Proc 3rd Int'l Workshop of Digital Mammography, Elsevier, New York, pp. 191-194, 1996.

29. Huo Z. Computerized methods for classification of masses and analysis of parenchynmal patterns on digitized mammograms. Ph.D. Dissertation, University of Chicago, June, 1998.

31. Amadasum M and King R. Texture features corresponding to texture properties. IEEE Trans on System, Man and Cybernetics 1989; 19:1264-1274.

32. Tahoces P, Correa J, Souto M, Gomes L and Vidal J. Computer-assisted diagnosis: The classification of mammographic breast parenchymal patterns. Phys Med Biol 1995; 40:103-117.

33. Jain AK. Fundamentals of Digital Image Processing. Englewood Cliffs, New Jersey, Prentice-Hall, 1986.

33. Katsuragawa S, Doi K, MacMahon H, Monnier-Cholley L, Ishida T and Kabayashi T. Classification of normal and abnormal lungs with interstitial disease by rule-based method and artificial neural networks. J Digit Imaging 1997; 10:108-114.

34. Caligiuri P, Giger ML, Favus MJ, Jia H, Doi K and Dixon LB. Computerized radiographic analysis of osteoporosis: Preliminary evaluation. Radiology 1993; 186:471-474.

35. Hays WL. Statistics. Philadelphia, Harcourt Brace College, 1994.

36. Moolgavkar SH, Prentice RL, eds. Modern Statistical Methods in Chronic Disease Epidemiology, John Wiley and Sons, pp. 50-62, 1986.

APPENDICES

1. Huo Z, Giger ML, Zhong W, Olopade OI: Analysis of relative contributions of mammographic features and age to breast cancer risk prediction. **Proceeding Paper** Digital Mammography 2000, Proc. 5th International Workshop on Digital Mammography, Medical Physics Publishing, Wisconsin pp. 732-736, 2001.

Analysis of Relative Contributions of Mammographic Features and Age to Breast Cancer Risk Prediction

ZHIMIN HUO
MARYELLEN L. GIGER
WEIMING ZHONG
Kurt Rossmann Laboratories for Radiologic Image Research, Department of Radiology
The University of Chicago, Chicago, Illinois

OLUFUNMILAYO I. OLOPADE
Department of Hematology and Oncology, The University of Chicago
Chicago, Illinois

INTRODUCTION

Studies based on visual assessment and computerized assessment of mammographic patterns showed that increasing mammographic density associated with an increased risk of breast cancer on the order of 4.0-6.0 between the most extensive and least extensive density patterns (Wolfe et al. 1987, Boyd et al. 1995, Byrne et al. 1995, Byng et al. 1997). Quantitative computerized analysis of mammographic patterns provides objective classification of density patterns, while the variability of visual assessment remains due to the subjective nature of human observers (Warner et al. 1992). We have developed computerized methods that characterize mammographic parenchymal patterns of women and relate these patterns to the risk of developing breast cancer. We have studied mammographic parenchymal patterns of cancer-free women who are at different risk levels of developing breast cancer, including BRCA1/BRCA2 mutation carriers, and women who have developed breast cancer (Huo et al. 2000a, b). A total of 14 mammographic features were extracted from the central breast region on digitized mammograms to characterize percent density of the breast or the heterogeneity/inhomogeneity (diffuse) patterns in the dense portions of the breast (Huo et al. 2000b). Three different approaches have been employed to relate these mammographic features to the risk of developing breast cancer (Huo et al. 2000a,b). In one approach, the features were related to risk as determined from existing clinical models (Gail and Claus models), which use well-known epidemiological factors such as a woman's age, her family history of breast cancer, reproductive history, etc. (Gail and Benichou 1992, Claus et al. 1993). Stepwise linear regression analysis was employed to identify useful features in predicting the risk as estimated from the Gail and Claus models. Four selected features, along with age, were used to predict the 10-year risks as estimated from the Gail and the Claus models. Results from linear regression analysis indicated that increasing age and mammographic density and coarse and low contrast mammographic patterns were positively correlated with increased breast cancer risk, yielding correlation coefficients (r) of 0.41 and 0.57 for the Gail and Claus models, respectively. Analysis of mammographic patterns of women who are BRCA1/BRCA2 mutation carriers and who have been diagnosed with breast cancer also suggested that women at

high risk of developing breast cancer tend to have dense breasts and their mammographic patterns tend to be coarse and low in contrast (Huo et al. 2000a,b).

It is important to understand the contribution of these computer-extracted mammographic features in predicting breast cancer risk and to study the potentials of these features in the prediction of breast cancer risk when they are incorporated with other risk factors into a model. The purpose of this study is to analyze the contribution of these computer-extracted mammographic features to breast cancer risk prediction in comparison with that of age, which is the most important single risk factor for breast cancer.

MATERIALS AND METHODS

Database

A total of 380 cancer-free cases were included in this study. Retrospective mammograms and information regarding the reproductive history, family history of breast cancer, and history of previous breast disease were collected for all cases to assess an individual's short-term risk (i.e., 10-year risk) of developing breast cancer. The 10-year risk is defined as the probability that a woman with given risk factors and given age will develop breast cancer in the next 10 years of her life. In this study, 10-year risks of developing breast cancer risk were estimated for all of the cases using both the Gail model and the Claus model. The 10-year risk was used for this study since the Claus model calculates short-term risk only up to the 10-year intervals. Mammograms from these cases were digitized using a Konica laser scanner (LD 4500; Konica Medical, Wayne, NJ) at 0.1 mm pixel size and 10-bit gray-level scale.

It should be noted that the cases used for the Gail and the Claus models were different since not all of the cases have complete information required by both the Gail and the Claus models. Of the 380 cases, 143 of them have the 10-year risk estimated from the Gail model and 303 of them have the 10-year risk as estimated from the Claus model.

Computer-extracted Mammographic Features

A total of 14 features were extracted from a region-of-interest (ROI) of size 256 pixels by 256 pixels, which was manually selected from the central region of the breast image. The central breast region was used because it usually includes the most dense parts of the breast. Detailed descriptions about these features can be found in the literature (Huo et al. 2000b). Useful features were then identified using the approach described above, i.e., stepwise linear regression analysis (Huo et al. 2000b), to predict the 10-year risk as estimated from the Gail or the Claus model. A total of four computer-extracted mammographic features, along with age, were selected from stepwise linear regression analysis. Age, skewness, coarseness, and contrast were selected for the 10-year risk as determined from the Gail model. Age, skewness, RMS variation, and coarseness were selected for the 10-year risk as determined from the Claus model. The skewness from gray-level histogram analysis and the root-mean-square (RMS) variation from the Fourier transform were

calculated to indicate the percent density. The coarseness and contrast features were obtained based on the spatial relationship among gray levels. They were used to characterize the heterogeneity of the dense tissue patterns within the ROIs.

Regression Analysis on Age and Mammographic Features

To assess the contribution of age alone in the prediction of an individual's 10-year risk, linear regression on age alone was performed. It should be noted that age is one of the risk factors used in both the Gail and the Claus models. The relationship between the 10-year risk and age was represented by a linear regression model. The proportion of the total variation in the 10-year risk explainable by age employing the linear regression model was used to quantify the "contribution" of age alone in the prediction of 10-year risk, as indicated by the squared correlation coefficient, r^2 (Hays 1994).

To assess the contribution of the selected mammographic features in the prediction of 10-year risk, linear regression on age and the mammographic features was performed to predict the 10-year risks as estimated from the Gail and the Clause models. The relationship of the 10-year risk with age and the selected mammographic features was represented by a multiple linear regression model. The "contribution" from age and the mammographic features together in the prediction of 10-year risk was indicated by the squared multiple correlation coefficient, r^2 (Hays 1994).

It should be noted that the r^2 ranges from 0 to 1, where $r^2 = 1$ indicates 100% of total variation in the observed values (e.g, 10-year risk as estimated from the Gail model) explained by the regression model or by the independent variables (e.g., the features). In other words, with $r^2 = 1$, all of the observed values for an individual's 10-year risk fall exactly on the straight "line" represented by the regression model.

Relative Contribution of Mammographic Features in Comparison with Age

Addition of any features to the regression model increases the squared multiple correlation coefficient, r^2. The increase in r^2 measures the additional worth of the added features but depends on the feature already in the model. The increase in r^2, when the mammographic features are added to the regression model, quantifes the percentage of the total variation in the 10-year risk explained by the mammographic features but not by age. As mentioned above, the contribution of age alone can be quantified in terms of r^2 when age is used alone. The additional contribution of mammographic features can be quantified in terms of the increase, Δr^2, in r^2 when these features are added. The relative contribution of these mammographic features in the prediction of an individual's 10-year risk is measured by the percent increase in r^2, i.e., $\Delta r^2/r^2$.

RESULTS

The models based on regression on age alone and mammographic features with age are listed in table 1 for the 10-year risk as estimated from the Gail and the Claus models. It should be noted that the analysis was performed separately

Table 1. Linear Regression Models on Age Alone and Along with Mammographic Features

10-year risk% (Gail model)	=	−0.025 + 0.001 age
10-year risk% (Gail model)	=	−0.03 − 0.004 skew + 34.51 cos −38.31 con + 0.002 age
10-year risk% (Claus model)	=	−0.076 + 0.003 age
10-year risk% (Claus model)	=	−0.09 − 0.013 skew + 0.002 rms −100.52 con + 0.004 age

NOTE: Skew, cos, con, and rms correspond to the skewness, coarse, contrast, and RMS variation.

for the risk as estimated from the Gail and the Claus models using the two different subsets of the database. As shown in table 1, the 10-year risk as estimated from the Gail model (303 cases) was positively correlated with age and coarseness, and was negatively correlated with skewness and contrast, yielding a correlation coefficient of 0.28 (p-value < 0.001) when age alone was used and a correlation coefficient of 0.41 (p-value < 0.001) when the mammographic features were included. The 10-year risk as estimated from the Claus model (143 cases) was positively correlated with age and RMS variation and was negatively correlated with skewness and contrast, yielding a correlation coefficient of 0.4 (p-value < 0.001) when age was used alone and a correlation coefficient of 0.57 (p-value < 0.001) when the mammographic features were added. The results imply that an individual's 10-year risk increases with age, with increasing mammographic density, and with coarse, low-contrast mammographic texture patterns.

In terms of contribution measured by r^2, regression on age alone yielded r^2s of 0.08 and 0.16 for the 10-year risks as estimated from the Gail and the Claus models, respectively. Regression on age and the selected mammographic features yielded r^2s of 0.17 and 0.32 for the 10-year risk as estimated from the Gail and the Claus models, respectively, which corresponds to increases of 113% and 100% in r^2.

DISCUSSION

Age has been identified as the most important risk predictor for breast cancer in women. Dense mammographic parenchymal patterns have been identified as one of the important risk factors for breast cancer. In this paper, we studied the association of the 10-year risks as estimated from the Gail and the Claus models with age and mammographic patterns as characterized by computer-extracted features using linear regression analysis. The contribution of age and the mammographic features to breast cancer risk prediction was quantified in terms of the squared correlation coefficient, r^2, i.e., the percentage of the total variation in the risk explainable by age alone or together with the mammographic features. The relative increases of 113% and 100% in r^2 for the 10-year risks as estimated from the Gail and the Claus models, respectively, indicate that the mammographic features, which were included in the regression model, contributed as

much as age in the prediction of breast cancer risk as estimated from the Gail and the Claus models, although the results need to be validated using a larger number of cases. Such a substantial contribution to the prediction of breast cancer risk, in comparison with that of age, indicates the importance of mammographic features in breast cancer risk prediction, and the need to incorporate them into a breast cancer risk prediction model.

ACKNOWLEDGMENTS

The authors thank Charles E. Metz, Ph.D., for his useful discussion. This work was supported in parts by USPHS grant CA79711 and a grant from The University of Chicago Cancer Research Center (NIH grant P20 CA66132), and grants from the US Army Medical Research and Materiel Command (DAMD 17-96-1-6058 and 17-99-1209).

Z. Huo and M. L. Giger are shareholders in R2 Technology, Inc. (Los Altos, CA). It is the policy of The University of Chicago that investigators disclose publicly actual or potential significant financial interests that may appear to be affected by the research activities.

REFERENCES

Boyd, N. F., J. Byng, and R. Jong (1995). "Quantitative classification of mammographic densities and breast cancer risk: Results from the Canadian National Breast Screening Study." *J. Natl. Cancer Inst.* 87: 670–675.

Byng, J. W., M. J. Yaffe, G. A. Lockwood, L. E. Little, D. L. Tritchler, and N. F. Boyd (1997). "Automated analysis of mammographic densities and breast carcinoma risk." *Cancer* 88: 66–74.

Byrne, C., C. Schairer, J. Wolfe, N. Parekh, M. Salane, L. A. Brinton, R. Hoover, and R. Haile (1995). "Mammographic features and breast cancer risk: Effects with time, age, and menopause status." *J. Natl. Cancer Inst.* 87: 1622–1629.

Claus, E. B., N. Risch, and W. D. Thompson (1993). "Autosomal dominant inheritance of early-onset breast cancer: Implications for risk prediction." *Cancer* 73: 643–651.

Gail, M. H., and J. Benichou. "Assessing the Risk of Breast Cancer in Individuals" in *Cancer Prevention*. V. T. DeVita, S. Hellman, and S. A. Rosenberg (eds.). Philadelphia: J. B. Lippincott, 1992.

Hays, W. L. *Statistics*. Philadelphia: Harcourt Brace College, 1994.

Huo, Z., M. L. Giger, and O. I. Olopade (2000a). "Computerized Analysis of Mammographic Patterns of Women With and Without Breast Cancer" in Proceedings of CARS'2000. San Francisco, 2000.

Huo, Z., M. L. Giger, D. E. Wolverton, W. Zhong, S. Cumming, and O. I. Olopade (2000b). "Computerized analysis of mammographic parenchymal patterns for breast cancer risk assessment: feature selection." *Med. Phys.* 27: 4–12.

Warner, E., G. Lockwood, M. Math, D. Tritchler, and N. F. Boyd (1992). "The risk of breast cancer associated with mammographic parenchymal patterns: A meta-analysis of the published literature to examine the effect of method of classification." *Cancer Detect. Prevent.* 16: 67–72.

Wolfe, J. N., A. F. Saftlas, and M. Salane (1987). "Mammographic parenchymal patterns and quantitative evaluation of mammographic densities: A case-control study." *Am. J. Roentgen.* 148: 1087–1092.